

Asymptotic Distribution Theory for Correspondence Analysis

CHARLES E. McCULLOCH*

The asymptotic distribution of the eigenvectors and eigenvalues in correspondence analysis is derived using a method of Anderson (1963). The results are illustrated on a condensed version of data of Maung (1941). The results are also applied to derive the asymptotic distribution of the eigenvectors and eigenvalues in principal components analysis of a correlation matrix from multivariate normal data.

KEY WORDS: Ordination techniques; Asymptotic distributions of eigenvectors and eigenvalues; Principal components analysis.

* Charles E. McCulloch is Assistant Professor, Biometrics Unit,
Cornell University, Ithaca, NY 14853. This paper is BU-786-M in the
Biometrics Unit Series, Cornell University, Ithaca, NY 14853.

1. INTRODUCTION

Correspondence analysis is a scaling technique which assigns a score to each of the categories of a cross-classification. Though it has come into widespread use in the ecology literature, e.g., Hill (1973), Swan (1970) or Whittaker (1967), it has received little statistical attention or investigation of its properties. Hill (1974) has demonstrated the equivalence between correspondence analysis of a two-way table of counts and canonical analysis of a contingency table and between correspondence analysis and principal components analysis. O'Neill (1978a, 1978b, 1981) has derived distributions for the canonical roots in contingency tables which thus may be used in some types of correspondence analysis. In correspondence analysis, attention focuses equally or more intently on the scores. Using work by Anderson (1963) on principal components analysis, the work of O'Neill is elaborated and extended. It is also shown how the same theorems can be used to handle the asymptotic distribution theory for principal components analysis performed on correlation matrices from a multivariate normal sample.

2. ASYMPTOTIC DISTRIBUTION THEORY

Anderson (1963, Theorem 1) derives the asymptotic distribution theory for the eigenvalues and eigenvectors of the sample variance-covariance matrix, $\hat{\Sigma}$, when the sample is from a multivariate normal distribution with variance-covariance matrix Σ . The proof can be divided into three parts:

1. Establishing asymptotic normality of the normalized elements of $\hat{\Sigma}$.
2. Establishing the functional convergence of the solutions of certain matrix equations as $\hat{\Sigma} \rightarrow \Sigma$.
3. Deriving conclusions about the distributions of the eigenvectors and eigenvalues of $\hat{\Sigma}$ using 1 and 2 above and Rubin's Theorem.

The key point is that many of the conclusions hold when the assumption of multivariate normality of the observations is replaced by the assumption that $n^{\frac{1}{2}}(\hat{\Sigma} - \Sigma)$ is asymptotically normal. Notably, part two of the proof remains intact. These conclusions are embodied in Theorem 1. The results are only stated for distinct roots, though, as in Anderson (1963), the results for sets of equal roots are straightforward.

Theorem 2 is a parallel theorem for the singular value decomposition of a sample matrix, as opposed to its spectral decomposition (as is used in principal components analysis).

Since we will be working with matrix distributions, we will make use of some special matrix results and it is convenient to introduce them now. The $\text{vec}(\cdot)$ operator is a matrix operator which stacks the columns of a matrix one upon the other to make a large column vector, i.e., the (i,j) element of A is the $n(i-1)+j$ element of the $mn \times 1$ column vector, $\text{vec } A$. A salient property of the $\text{vec}(\cdot)$ operator is that

$$\text{vec}(ABC) = (C' \otimes A) \text{vec} B \quad ,$$

where \otimes denotes the Kronecker product. A special matrix we will make use of, connected with the above, is the commutation matrix, $K_{m,n}$ (Magnus and Neudecker, 1979). We will use the following property of $K_{m,n}$,

$$\text{vec} \begin{pmatrix} A \\ \end{pmatrix}_{m \times n} = K_{m,n} \text{vec} \begin{pmatrix} A' \\ \end{pmatrix}_{n \times m} \quad .$$

For descriptions of these and other results see Henderson and Searle (1979).

Suppose that Σ is positive definite with spectral decomposition given by

$$\begin{matrix} \Sigma \\ p \times p \end{matrix} = \Gamma \Lambda \Gamma' \quad ,$$

where $\Gamma\Gamma' = \Gamma'\Gamma = I_p$ and $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\}$ $\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$. We assume that $\gamma_{ii} > 0$ so that Γ is uniquely defined. Write a similar decomposition for $\hat{\Sigma}$, which is assumed to be positive definite with probability one,

$$\hat{\Sigma}_{p \times p} = \hat{\Gamma} \hat{\Lambda} \hat{\Gamma}'$$

Theorem 1. Write $\Gamma = (\gamma_1, \gamma_2, \dots, \gamma_p)$. Notation as above. Assume $n^{\frac{1}{2}}(\text{vec } \hat{\Sigma} - \text{vec } \Sigma) \sim AN(0, V)$, then

$$n^{\frac{1}{2}}(\hat{\lambda} - \lambda) \sim AN(0, W_{\lambda}) \quad (2.1)$$

$$n^{\frac{1}{2}}(\hat{\gamma}_k - \gamma_k) \sim AN(0, W_k) \quad k = 1, 2, \dots, p \quad (2.2)$$

$$\text{asy. cov}[n^{\frac{1}{2}}(\hat{\gamma}_k - \gamma_k), n^{\frac{1}{2}}(\hat{\gamma}_j - \gamma_j)] = \sum_{\ell \neq k} \sum_{m \neq j} \gamma_{\ell} \gamma_m' \frac{(\gamma_{\ell}' \otimes \gamma_k') V (\gamma_m \otimes \gamma_j)}{(\lambda_k - \lambda_{\ell})(\lambda_j - \lambda_m)}, \quad (2.3)$$

where $(W_{\lambda})_{jk} = [(\gamma_j' \otimes \gamma_j') V (\gamma_k \otimes \gamma_k)]_{jk}$ and W_k is specified by (2.3).

Proof. See the Appendix.

Similar results can be derived for the singular values of a sample matrix. As we will see in Section 3, the distributions of the decomposing matrices can be found using Theorem 1, so we will concentrate on the sample singular values. Suppose that $\hat{B}_{p \times q}$ ($p \leq q$) is an estimator of B such that

$$n^{\frac{1}{2}}(\text{vec } \hat{B} - \text{vec } B) \sim AN(0, Z)$$

Suppose also that B has singular value decomposition (Seber, 1977, p. 393) given by

$$B = S(\Sigma \ 0)T'$$

where

$$\begin{aligned} SS' &= S'S = I_p, & s_{ii} &> 0, \\ TT' &= T'T = I_q, \\ \Sigma &= \text{diag}\{\sigma_1, \sigma_2, \dots, \sigma_p\}, & \sigma_1 &> \sigma_2 > \dots > \sigma_p > 0. \end{aligned} \quad (2.4)$$

This decomposition is unique, except perhaps for the last $p - q$ columns of T ($t_{p+1}, t_{p+2}, \dots, t_q$) and has the property that

$$t_i = \sigma_i^{-1} B' s_i \quad i = 1, 2, \dots, p \quad .$$

Theorem 2. Write $S = (s_1, \dots, s_p)$ and $T = (t_1, \dots, t_q)$. Notation as above. Under the above assumptions,

$$n^{\frac{1}{2}}(\hat{g} - g) \sim AN(0, W_g) \quad ,$$

where

$$(W_g)_{jk} = [(t_j' \otimes s_j') Z(t_k \otimes s_k)]_{jk} \quad .$$

Proof. See the Appendix.

The utility of these two theorems is that they translate the statements about the asymptotic normality of sample matrices into asymptotic normality of decomposing matrices.

3. ASYMPTOTIC DISTRIBUTIONS FOR CORRESPONDENCE ANALYSIS

The simplest form of correspondence analysis, that of a two-way contingency table of counts, is equivalent (Hill, 1974) to a canonical analysis of the contingency table. The canonical vectors are interpreted as scores in an ordination procedure. That is, the scores are used to order the row categories and the column categories. The ordinations are then interpreted as a gradient, hopefully making sense in the problem at hand.

Let n_{ij} be the count in the (i, j) cell of a contingency table ($i = 1, 2, \dots, p$, $j = 1, 2, \dots, q$, $p \leq q$) and assume that the counts are multinomially distributed with $E[n_{ij}] = n \cdot p_{ij}$. In the usual notation, $n_{i.}, n_{.j}$ and $p_{i.}, p_{.j}$ denote the marginal totals and marginal probabilities. Let $\hat{B} = [n_{ij} / (n_{i.} n_{.j})^{\frac{1}{2}}]$ and suppose it has singular value decomposition as in (2.4),

$$\hat{B} = \hat{S}(\hat{\Sigma} \ 0)\hat{T}',$$

where

$$\hat{S} = (\hat{s}_{\sim 1}, \hat{s}_{\sim 2}, \dots, \hat{s}_{\sim p}) \quad \text{and} \quad \hat{T} = (\hat{t}_{\sim 1}, \hat{t}_{\sim 2}, \dots, \hat{t}_{\sim q}).$$

Sets of scores are then found from $(\hat{s}_{\sim i}, \hat{t}_{\sim i})$ $i=1, 2, \dots, p$ and are used for ordination. These sets of scores are ordered in importance by their singular values, $\hat{\sigma}_{\sim i}$, just as sets of canonical vectors are ordered by eigenvalues in a canonical correlation analysis. The ties between correspondence analysis and canonical analysis of contingency tables are that

1. The usual χ^2 test for independence is related to the singular values

$$\chi^2 = \sum_{i=1}^p \sum_{j=1}^q (n_{ij} - n_{i.}n_{.j}/n_{..})^2 / (n_{i.}n_{.j}/n_{..}) = n_{..} \sum_{i=2}^p \hat{\sigma}_{\sim i}^2.$$

2. The canonical vectors or canonical polynomials $(\hat{\xi}_{\sim i}^*, \hat{\eta}_{\sim j}^*, \text{ O'Neill, 1978b})$ are related to the decomposing matrices \hat{S} and \hat{T} :

$$\begin{aligned} \hat{s}_{\sim i} &= \text{diag}\{n_{\sim 1.}^{-\frac{1}{2}}, n_{\sim 2.}^{-\frac{1}{2}}, \dots, n_{\sim p.}^{-\frac{1}{2}}\} \hat{\xi}_{\sim i}^* \\ \hat{t}_{\sim j} &= \text{diag}\{n_{.1}^{-\frac{1}{2}}, n_{.2}^{-\frac{1}{2}}, \dots, n_{.q}^{-\frac{1}{2}}\} \hat{\eta}_{\sim j}^* \end{aligned}$$

The asymptotic distribution theory for correspondence analysis can be obtained by applying Theorem 1 to $\hat{B}\hat{B}'$ and Theorem 2 to \hat{B} . Let us first consider the sample singular values, $\hat{\sigma}_{\sim i}$.

Lemma 1. Let $\hat{B} = [n_{ij}/(n_{i.}n_{.j})^{\frac{1}{2}}] = [\hat{p}_{ij}/(\hat{p}_{i.}\hat{p}_{.j})^{\frac{1}{2}}]$. Then

$$n^{\frac{1}{2}}(\text{vec } \hat{B} - \text{vec } B) \sim AN(0, Z),$$

where

$$\begin{aligned} V &= \text{diag}\{\text{vec } B\}(\text{diag}^{-1}\{\text{vec } P\} - \frac{3}{4} \mathbf{1}_{\sim q} \mathbf{1}_{\sim q}' \otimes R^{-1} \\ &\quad - \frac{3}{4} C^{-1} \otimes \mathbf{1}_{\sim p} \mathbf{1}_{\sim p}' + \frac{1}{4} \mathbf{1}_{\sim q} \otimes Q \otimes \mathbf{1}_{\sim p}' + \frac{1}{4} \mathbf{1}_{\sim q}' \otimes Q \otimes \mathbf{1}_{\sim p}) \text{diag}\{\text{vec } B\}, \end{aligned} \quad (3.1)$$

and $P = (p_{ij})$, $R = \text{diag}\{p_{i.}\}$, $C = \text{diag}\{p_{.j}\}$, $Q = (p_{ij}/p_{i.}p_{.j})$ and $\mathbf{1}'_n = (1 \ 1 \ \dots \ 1)_{1 \times n}$.

Proof. It is convenient to first work with $\ln \hat{b}_{kl}$. Using a theorem of Bishop, Fienberg and Holland (1975, Theorem 14.6-4), $n^{\frac{1}{2}}(\ln \hat{b}_{kl} - \ln b_{kl})$ is asymptotically multivariate normal with means zero and covariances given by

$$\begin{aligned} \text{asy. cov}(n^{\frac{1}{2}} \ln \hat{b}_{kl}, n^{\frac{1}{2}} \ln \hat{b}_{k'l'}) \\ &= \sum_g \sum_{g'} \sum_h \sum_{h'} \left(\frac{\partial \ln b_{kl}}{\partial p_{gh}} \right) \left(\frac{\partial \ln b_{k'l'}}{\partial p_{g'h'}} \right) \text{cov}(\hat{p}_{gh}, \hat{p}_{g'h'}) \\ &= \sum_g \sum_h \left(\frac{\partial \ln b_{kl}}{\partial p_{gh}} \right) \left(\frac{\partial \ln b_{k'l'}}{\partial p_{gh}} \right) p_{gh} - \left[\sum_g \sum_h \frac{\partial \ln b_{kl}}{\partial p_{gh}} p_{gh} \right] \times \left[\sum_{g'} \sum_{h'} \frac{\partial \ln b_{k'l'}}{\partial p_{g'h'}} p_{g'h'} \right]. \end{aligned}$$

However,

$$\begin{aligned} \sum_g \sum_h \left(\frac{\partial \ln b_{kl}}{\partial p_{gh}} \right) p_{gh} &= \sum_g \sum_h \left(\frac{\delta_{kg} \delta_{lh}}{p_{kl}} - \frac{1}{2} \frac{\delta_{kg}}{p_{k.}} \frac{1}{2} \frac{\delta_{lh}}{p_{.l}} \right) p_{gh} \\ &= \frac{p_{kl}}{p_{kl}} - \frac{1}{2} \frac{p_{k.}}{p_{k.}} - \frac{1}{2} \frac{p_{.l}}{p_{.l}} = 0, \end{aligned}$$

(δ_{ij} is the Kronecker delta)

so that $\text{asy. cov}(n^{\frac{1}{2}} \ln \hat{b}_{kl}, n^{\frac{1}{2}} \ln \hat{b}_{k'l'})$ is given by

$$\begin{aligned} \sum_g \sum_h \left(\frac{\partial \ln b_{kl}}{\partial p_{gh}} \right) \left(\frac{\partial \ln b_{k'l'}}{\partial p_{gh}} \right) p_{gh} \\ &= \sum_g \sum_h \left(\frac{\delta_{kg} \delta_{lh}}{p_{kl}} - \frac{1}{2} \frac{\delta_{kg}}{p_{k.}} - \frac{1}{2} \frac{\delta_{lh}}{p_{.l}} \right) \times \left(\frac{\delta_{k'g} \delta_{l'h}}{p_{k'l'}} - \frac{1}{2} \frac{\delta_{k'g}}{p_{k'.}} - \frac{1}{2} \frac{\delta_{l'h}}{p_{.l'}} \right) p_{gh} \\ &= \sum_g \sum_h \left(\frac{\delta_{kg} \delta_{k'g} \delta_{lh} \delta_{l'h}}{p_{kl} p_{k'l'}} - \frac{1}{2} \frac{\delta_{kg} \delta_{lh} \delta_{k'g}}{p_{kl} p_{k'.}} - \frac{1}{2} \frac{\delta_{kg} \delta_{lh} \delta_{l'h}}{p_{kl} p_{.l'}} \right. \\ &\quad \left. - \frac{1}{2} \frac{\delta_{kg} \delta_{k'g} \delta_{l'h}}{p_{k.} p_{k'l'}} + \frac{1}{4} \frac{\delta_{kg} \delta_{k'g}}{p_{k.} p_{k'.}} + \frac{1}{4} \frac{\delta_{kg} \delta_{l'h}}{p_{k.} p_{.l'}} \right) p_{gh} \end{aligned}$$

$$\begin{aligned}
 & - \frac{1}{2} \frac{\delta_{lh} \delta_{k'g} \delta_{l'h}}{p_{.l} p_{k'.l}} + \frac{1}{4} \frac{\delta_{lh} \delta_{k'g}}{p_{.l} p_{k'..}} + \frac{1}{4} \frac{\delta_{lh} \delta_{l'h}}{p_{.l} p_{.l.}} \Big) p_{gh} \\
 & = \frac{\delta_{kk'} \delta_{ll'}}{p_{kl}} - \frac{3}{4} \frac{\delta_{kk'}}{p_{k.}} - \frac{3}{4} \frac{\delta_{ll'}}{p_{.l}} + \frac{1}{4} \frac{p_{kl'}}{p_{k.p.l}} + \frac{1}{4} \frac{p_{k'l}}{p_{k'.p.l}} .
 \end{aligned}$$

Using the δ -method (Bishop, Fienberg and Holland, 1975, Theorem 14.6-2) to go from the distribution of $n^{\frac{1}{2}} \ln \hat{b}_{kl}$ to $n^{\frac{1}{2}} \hat{b}_{kl}$ yields the result that $n^{\frac{1}{2}}(\hat{b}_{kl} - b_{kl})$ are asymptotically multivariate normal with means zero and covariances given by

$$\begin{aligned}
 & \text{asy. cov}(n^{\frac{1}{2}} \hat{b}_{kl}, n^{\frac{1}{2}} \hat{b}_{k'l'}) \\
 & = b_{kl} b_{k'l'} \left(\frac{\delta_{kk'} \delta_{ll'}}{p_{kl}} - \frac{3}{4} \frac{\delta_{kk'}}{p_{k.}} - \frac{3}{4} \frac{\delta_{ll'}}{p_{.l}} + \frac{1}{4} \frac{p_{kl'}}{p_{k.p.l}} + \frac{1}{4} \frac{p_{k'l}}{p_{k'.p.l}} \right) .
 \end{aligned}$$

This can be written in matrix notation as

$$n^{\frac{1}{2}}(\text{vec} \hat{B} - \text{vec} B) \sim \text{AN}(0, Z) ,$$

where

$$\begin{aligned}
 Z = \text{diag}\{\text{vec } B\} & \left[\text{diag}^{-1}\{\text{vec}(p_{ij})\} - \frac{3}{4} \mathbf{1}_{\tilde{q}} \mathbf{1}_{\tilde{q}}' \otimes \text{diag}^{-1}\{p_{i.}\} \right. \\
 & - \frac{3}{4} \text{diag}^{-1}\{p_{.j}\} \mathbf{1}_{\tilde{p}} \mathbf{1}_{\tilde{p}}' + \frac{1}{4} \mathbf{1}_{\tilde{q}} \otimes \left(\frac{p_{ij}}{p_{i.p.j}} \right) \otimes \mathbf{1}_{\tilde{p}}' \\
 & \left. + \frac{1}{4} \mathbf{1}_{\tilde{q}}' \otimes \left(\frac{p_{ji}}{p_{j.p.i}} \right) \otimes \mathbf{1}_{\tilde{p}} \right] \text{diag}\{\text{vec } B\} .
 \end{aligned}$$

Theorem 3. Let the matrix $B = [p_{ij}/(p_{i.p.j})^{\frac{1}{2}}]$ have singular value decomposition,

$$B = S(\Sigma \ 0)T'$$

as in (2.4), with a similar decomposition for $\hat{B} = [\hat{p}_{ij}/(\hat{p}_{i.p.j})^{\frac{1}{2}}]$,

$$\hat{B} = \hat{S}(\hat{\Sigma} \ 0)\hat{T}' ,$$

then

$$n^{\frac{1}{2}}(\hat{\sigma} - \sigma) \sim AN(0, W_{\sigma}^{\wedge})$$

where

$$(W_{\sigma}^{\wedge})_{jk} = (t_j' \otimes s_j') Z(t_k \otimes s_k)$$

and Z is given by (3.1).

Proof. Applying Theorem 2 to \hat{B} gives the desired result.

Calculating the asymptotic distribution as in Theorem 3 obviates the need of specifying particular preliminary transformations, as in O'Neill (1978a,b). It also represents the results in matrix notation.

Although O'Neill (1978b, equation 19) calculates an asymptotic expansion for the vectors \hat{s}_i , he never explicitly calculates the limiting normal distribution. With the use of Theorem 1 we now calculate the limit distribution of \hat{s}_i and \hat{t}_j .

Since $B = S(\Sigma \ 0)T'$, we have

$$BB' = S\Sigma^2S'$$

and therefore S is the decomposing matrix in the spectral decomposition of BB' . To get the asymptotic distribution of \hat{s}_i we may apply Theorem 1 to \hat{BB}' . First we need a lemma about the asymptotic normality of \hat{BB}' .

Lemma 2. Notation as above.

$$n^{\frac{1}{2}}(\text{vec } S'\hat{BB}'S - \text{vec } \Sigma^2) \sim AN(0, V^*) ,$$

where

$$V^* = (I_{p^2} + K_{p,p})(S' \otimes \Sigma T_1') Z(S \otimes T_1 \Sigma)(I_{p^2} + K_{p,p}) ,$$

and

$$T_1 = (t_1, t_2, \dots, t_p) .$$

Proof. First note that $(S'\hat{B}\hat{B}'S)_{ij} = s_i'\hat{B}\hat{B}'s_j$. Working from the identity

$$n^{\frac{1}{2}}s_i'(\hat{B}\hat{B}' - BB')s_j = n^{\frac{1}{2}}s_i'B(\hat{B} - B)'s_j + n^{\frac{1}{2}}s_i'(\hat{B} - B)B's_j + n^{\frac{1}{2}}s_i'(\hat{B} - B)(\hat{B} - B)'s_j$$

we obtain

$$n^{\frac{1}{2}}(s_i'\hat{B}\hat{B}'s_j - \sigma_i\sigma_j\delta_{ij}) = n^{\frac{1}{2}}s_i'B(\hat{B} - B)'s_j + n^{\frac{1}{2}}s_i'(\hat{B} - B)B's_j + o_p(1) \quad (3.2)$$

by Lemma 1. Also by Lemma 1, (3.2) shows that $n^{\frac{1}{2}}(\text{vec } S'\hat{B}\hat{B}'S - \text{vec } \Sigma^2)$ is asymptotically normal with variance-covariance matrix given by

$$\begin{aligned} V^* &= \text{asy. var}[\text{vec } S'Bn^{\frac{1}{2}}(\hat{B} - B)'S + \text{vec } S'n^{\frac{1}{2}}(\hat{B} - B)B'S] \\ &= \text{asy. var}[(K_{p,p} + I_{p^2})\text{vec } S'Bn^{\frac{1}{2}}(\hat{B} - B)S] \\ &= (I_{p^2} + K_{p,p})(S' \otimes S'B)\text{asy. var}[n^{\frac{1}{2}}\text{vec}(\hat{B} - B)](S \otimes B'S)(I_{p^2} + K_{p,p}) \\ &= (I_{p^2} + K_{p,p})(S' \otimes \Sigma T_1')Z(S \otimes T_1'\Sigma)(I_{p^2} + K_{p,p}) \quad , \end{aligned}$$

where T_1 is a matrix consisting of the first p columns of T .

Remark. If we write $u_{ij} = n^{\frac{1}{2}}[s_i'(\hat{B} - B)t_j]$, then another way to represent the results of this lemma is $n^{\frac{1}{2}}(s_i'\hat{B}\hat{B}'s_j - \sigma_i\sigma_j\delta_{ij}) = \sigma_i u_{ij} + \sigma_j u_{ji} + o_p(1)$. We are now ready to derive the asymptotic distributions of the \hat{s}_i and \hat{t}_j .

Theorem 4. Notation and assumptions as in Theorem 3, and let

$$w_{abcd} = \text{asy. cov}[s_a'(\hat{B} - B)t_b, s_c'(\hat{B} - B)t_d] \quad ,$$

then

$$\begin{aligned} n^{\frac{1}{2}}(\hat{s}_i - s_i) & \quad i = 1, 2, \dots, p \quad , \\ n^{\frac{1}{2}}(\hat{t}_j - t_j) & \quad j = 1, 2, \dots, p \quad , \end{aligned}$$

are asymptotically multivariate normal with means zero and variances and covariances given by

$$\begin{aligned} & \text{asy. cov}[n^{\frac{1}{2}}(\hat{s}_i - s_i), n^{\frac{1}{2}}(\hat{s}_k - s_k)] \\ &= \sum_{l \neq i} \sum_{m \neq k} s_l s'_m \frac{\sigma_i \sigma_m w_{ilmk} + \sigma_i \sigma_k w_{iklm} + \sigma_l \sigma_m w_{limk} + \sigma_l \sigma_k w_{likm}}{(\sigma_i^2 - \sigma_l^2)(\sigma_k^2 - \sigma_m^2)} \end{aligned} \quad (3.3)$$

$$= \sum_{l \neq i} \sum_{m \neq k} s_l s'_m \frac{V^*_{(l-1)p+i, (m-1)p+k}}{(\sigma_i^2 - \sigma_l^2)(\sigma_k^2 - \sigma_m^2)} \quad i, k = 1, 2, \dots, p \quad (3.4)$$

$$\begin{aligned} & \text{asy. cov}[n^{\frac{1}{2}}(\hat{t}_j - t_j), n^{\frac{1}{2}}(\hat{t}_n - t_n)] \\ &= \sum_{\substack{l=1 \\ l \neq j}}^p \sum_{\substack{m=1 \\ m \neq n}}^p t_l t'_m \frac{\sigma_j \sigma_m w_{jlmn} + \sigma_j \sigma_n w_{jlnm} + \sigma_l \sigma_m w_{ljmn} + \sigma_l \sigma_n w_{ljnm}}{(\sigma_j^2 - \sigma_l^2)(\sigma_n^2 - \sigma_m^2)} \\ &+ \sum_{\substack{l=1 \\ l \neq j}}^p \sum_{\substack{m=p+1 \\ m \neq n}}^q t_l t'_m \frac{\sigma_j w_{jlnm} + \sigma_l w_{ljnm}}{(\sigma_j^2 - \sigma_l^2) \sigma_n} \\ &+ \sum_{\substack{l=p+1 \\ l \neq j}}^q \sum_{\substack{m=1 \\ m \neq n}}^p t_l t'_m \frac{\sigma_n w_{jlnm} + \sigma_m w_{jlmn}}{(\sigma_n^2 - \sigma_m^2) \sigma_j} \\ &+ \sum_{l=p+1}^q \sum_{m=p+1}^q t_l t'_m \frac{w_{jlnm}}{\sigma_j \sigma_n} . \end{aligned} \quad (3.5)$$

Proof. To derive the asymptotic distribution of the \hat{s}_i we can apply Lemma 2 and Theorem 1 to conclude that

$$n^{\frac{1}{2}}(\hat{s}_i - s_i) \sim AN(0, W_k) ,$$

with variances and covariances given by

$$\text{asy. cov}[n^{\frac{1}{2}}(\hat{s}_i - s_i), n^{\frac{1}{2}}(\hat{s}_k - s_k)] = \sum_{l \neq i} \sum_{m \neq k} s_l s'_m \frac{(s'_l \otimes s'_i) V(s_m \otimes s_k)}{(\lambda_i - \lambda_l)(\lambda_k - \lambda_m)} .$$

For this case, $\lambda_j = \sigma_j^2$ and $(s'_l \otimes s'_i) V(s_m \otimes s_k)$ is given by Lemma 2 as

$V_{(\ell-1)p+i, (m-1)p+k}^*$ or following the Remark, as

$$\sigma_i \sigma_m w_{ilmk} + \sigma_i \sigma_k w_{ilk m} + \sigma_\ell \sigma_m w_{limk} + \sigma_\ell \sigma_k w_{likm} \quad .$$

This completes the proof for the $\hat{s}_{\sim j}$. To derive the distribution of $\hat{t}_{\sim j}$ ($j \leq p$) we use the relation

$$\hat{t}_{\sim j} = \hat{\sigma}_j^{-1} \hat{B}' \hat{s}_{\sim j} \quad .$$

Arguing as in Lemma 2 shows that

$$\begin{aligned} n^{\frac{1}{2}}(\hat{t}_{\sim j} - t_{\sim j}) &= B' s_{\sim j} n^{\frac{1}{2}}(\hat{\sigma}_j^{-1} - \sigma_j^{-1}) + B' n^{\frac{1}{2}}(\hat{s}_{\sim j} - s_{\sim j}) \sigma_j^{-1} \\ &\quad + n^{\frac{1}{2}}(\hat{B} - B)' s_{\sim j} \sigma_j^{-1} + o_p(1) \quad . \end{aligned} \quad (3.6)$$

Let us simplify (3.6) term by term. Using Theorem 2 and the δ -method shows that $n^{\frac{1}{2}}(\hat{\sigma}_j^{-1} - \sigma_j^{-1}) = - (u_{jj}/\sigma_j^2) + o_p(1)$. Since $B' s_{\sim j} = \sigma_j t_{\sim j}$, the first term simplifies to $-t_{\sim j}(u_{jj}/\sigma_j) + o_p(1)$. The second term can be represented using the first part of the proof as

$$\begin{aligned} B' \left(\sum_{\substack{\ell=1 \\ \ell \neq j}}^p s_{\sim \ell} \frac{\sigma_j \sigma_\ell + \sigma_\ell u_{j\ell}}{(\sigma_j^2 - \sigma_\ell^2)} \right) \sigma_j^{-1} + o_p(1) \\ = \sum_{\substack{\ell=1 \\ \ell \neq j}}^p \sigma_j^{-1} \sigma_\ell t_{\sim \ell} \frac{\sigma_j u_{j\ell} + \sigma_\ell u_{\ell j}}{(\sigma_j^2 - \sigma_\ell^2)} + o_p(1) \quad . \end{aligned}$$

The third term can be rewritten as

$$n^{\frac{1}{2}}(\hat{B} - B) s_{\sim j} \sigma_j^{-1} = \left(\sum_{\ell=1}^q t_{\sim \ell} t'_{\sim \ell} \right) n^{\frac{1}{2}}(\hat{B} - B) s_{\sim j} \sigma_j^{-1} = \sum_{\ell=1}^q t_{\sim \ell} u_{j\ell} \sigma_j^{-1} \quad .$$

Combining the results and simplifying yields the result,

$$n^{\frac{1}{2}}(\hat{t}_{\sim j} - t_{\sim j}) = \sum_{\substack{\ell=1 \\ \ell \neq j}}^p t_{\sim \ell} \frac{u_{j\ell} \sigma_j + u_{\ell j} \sigma_\ell}{(\sigma_j^2 - \sigma_\ell^2)} + \sum_{\ell=p+1}^q t_{\sim \ell} \frac{u_{j\ell}}{\sigma_j} + o_p(1) \quad .$$

The remainder of the proof follows.

Remark. As it appears in (3.5) it seems that $\text{asy. var}[n^{\frac{1}{2}}(\hat{t}_j - t_j)]$ cannot be consistently estimated since it depends on t_j with $j > p$. The representation given in (3.6) shows this is not the case.

4. AN EXAMPLE

We illustrate the previous computations with a condensed version of an example due to Maung (1941) and used in Lancaster (1958), reproduced in Table 1.

-TABLE 1 HERE-

The data are a cross-classification of school children by eye and hair color. The matrix

$$\hat{B}_{2 \times 3} = \left(\frac{\hat{p}_{ij}}{(\hat{p}_{i.} \hat{p}_{.j})^{\frac{1}{2}}} \right) = \begin{bmatrix} .534866 & .635599 & .321250 \\ .082947 & .282126 & .476424 \end{bmatrix}$$

and has singular value decomposition,

$$\hat{B} = \hat{S}(\hat{\Sigma} \ 0) \hat{T}'$$

$$= \begin{bmatrix} .876681 & .481072 \\ .481072 & -.876681 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & .32669 & 0 \end{bmatrix} \begin{bmatrix} .508810 & .565038 & .649495 \\ .692941 & .178869 & -.698455 \\ .510828 & -.805443 & .300528 \end{bmatrix}' .$$

The interpretation of the second column of T is that the hair categories are assigned scores of .565, .179 and -.805 and are ordered as expected from dark to light.

Also,

$$\text{asy. cov}[n^{\frac{1}{2}}(\text{vec } \hat{B} - \text{vec } B)] = \hat{Z}$$

$$= \begin{array}{c} \begin{array}{cccccc} & 11 & 21 & 12 & 22 & 13 & 23 \\ 11 & .262645 & -.111469 & -.140913 & .052254 & -.085351 & .067287 \\ 21 & & .297809 & .026959 & -.068904 & .020903 & -.105565 \\ 12 & & & .232380 & -.195244 & -.109226 & .118375 \\ 22 & & & & .497738 & .067634 & -.342006 \\ 13 & & & & & .357031 & -.338267 \\ 23 & & & & & & .770758 \end{array} \end{array}$$

and $(\hat{T}' \otimes \hat{S}') \hat{Z} (\hat{T} \otimes \hat{S})$ is given by

$$= \begin{array}{c} \begin{array}{cccccc} & 11 & 21 & 12 & 22 & 13 & 23 \\ 11 & 0 & 0 & 0 & 0 & 0 & 0 \\ 21 & & .223319 & .072956 & -.079895 & 0 & -.040394 \\ 12 & & & .223319 & -.008299 & 0 & -.068420 \\ 22 & & & & .971995 & -.061822 & -.310744 \\ 13 & & & & & .250000 & -.036913 \\ 23 & & & & & & .749728 \end{array} \end{array}$$

The numbers across the top and sides are to indicate which entries are referred to in the original matrix. From this $\hat{W}_{\hat{\sigma}} = \text{asy. cov}[n^{\frac{1}{2}}(\hat{\sigma} - \sigma)]$ can be read off as

$$\hat{W}_{\hat{\sigma}} = \begin{bmatrix} 0 & 0 \\ 0 & .971995 \end{bmatrix}.$$

The zero entries are to be expected since $\hat{\sigma}_1 \equiv 1$. From this an approximate confidence interval for σ_2 can be found as

$$\hat{\sigma}_2 \pm 2\text{s.e.}(\hat{\sigma}_2) = .32669 \pm 2(.971995/22,361)^{\frac{1}{2}} = (.313504, .339876).$$

so the result is

$$\text{asy. var}[n^{\frac{1}{2}}(\hat{t}_2 - t_2)] = \begin{bmatrix} 1.21768 & -.92644 & .64849 \\ & .96888 & -.43476 \\ & & .35838 \end{bmatrix} .$$

This could be used to set individual asymptotic confidence intervals on the \hat{t}_{k2} or to set a confidence ellipsoid on \hat{t}_2 .

5. ASYMPTOTIC DISTRIBUTIONS FOR PRINCIPAL COMPONENTS ANALYSIS BASED ON A CORRELATION MATRIX

Theorem 1 can also be used to derive the asymptotic distribution of the eigenvectors and eigenvalues of the sample correlation matrix. Let $R = (r_{ij}) = [\hat{\sigma}_{ij}/(\hat{\sigma}_{ii}\hat{\sigma}_{jj})^{\frac{1}{2}}]$ and $\bar{R} = [\sigma_{ij}/(\sigma_{ii}\sigma_{jj})^{\frac{1}{2}}]$. As in Section 3, the asymptotic normality of $\hat{\Sigma}$ implies the asymptotic normality of R .

Lemma 4. Let R be the sample correlation matrix from a random sample of size n from a multivariate normal distribution with positive definite variance-covariance matrix, Σ . Then

$$n^{\frac{1}{2}}(\text{vec } R - \text{vec } \bar{R}) \sim AN(0, V_R) ,$$

where V_R has entries given by

$$\begin{aligned} \text{asy. cov}(n^{\frac{1}{2}}r_{ij}, n^{\frac{1}{2}}r_{i'j'}) &= \bar{r}_{ij}\bar{r}_{i'j'} \left\{ \frac{\sigma_{ij}\sigma_{ji} + \sigma_{ii}\sigma_{jj}}{\sigma_{ij}\sigma_{i'j'}} \right. \\ &\quad - \left[\frac{\sigma_{ii}\sigma_{ji}}{\sigma_{ij}\sigma_{i'i'}} + \frac{\sigma_{ij}\sigma_{jj}}{\sigma_{ij}\sigma_{j'j'}} + \frac{\sigma_{ij}\sigma_{ii}}{\sigma_{ii}\sigma_{i'j'}} + \frac{\sigma_{jj}\sigma_{ji}}{\sigma_{jj}\sigma_{i'j'}} \right] \\ &\quad \left. + \frac{1}{2}[\bar{r}_{ii}^2 + \bar{r}_{jj}^2 + \bar{r}_{ij}^2 + \bar{r}_{ji}^2] \right\} . \end{aligned}$$

To get confidence intervals for the \hat{s}_i or \hat{t}_j we need asy. cov of $\text{vec } \hat{B}\hat{B}'$.

We will illustrate for the \hat{t}_j .

The asy. $\hat{\text{var}} \left[n^{\frac{1}{2}} \text{vec}(\hat{T}'\hat{B}'\hat{B}\hat{T}) - \text{vec} \begin{pmatrix} \Sigma^2 & 0 \\ 0 & 0 \end{pmatrix} \right]$ is given by

$$\begin{array}{c} \begin{matrix} & 11 & 21 & 31 & 12 & 22 & 32 & 13 & 23 & 33 \end{matrix} \\ \begin{matrix} 11 \\ 21 \\ 31 \\ 12 \\ 22 \\ 32 \\ 13 \\ 23 \\ 33 \end{matrix} \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ .1995 & 0 & .1995 & -.0116 & .0180 & 0 & .0180 & 0 & 0 \\ .2500 & 0 & .0404 & .0121 & .2500 & .0121 & 0 & 0 \\ .1995 & -.0116 & .0180 & 0 & .6180 & 0 & 0 & 0 \\ .4149 & -.0663 & .0404 & -.0663 & 0 & 0 & 0 & 0 \\ .0800 & .0121 & .0800 & 0 & 0 & 0 & 0 & 0 \\ .2500 & .0121 & 0 & 0 & 0 & 0 & 0 & 0 \\ .0800 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \end{array} .$$

From this we can obtain, for example,

$$\begin{aligned} \text{asy. } \hat{\text{var}}[n^{\frac{1}{2}}(\hat{t}_2 - t_2)] &= \hat{t}_1' \hat{t}_1' \frac{\hat{v}_{4,4}^*}{(\hat{\sigma}_2^2 - \hat{\sigma}_1^2)(\hat{\sigma}_2^2 - \hat{\sigma}_1^2)} + \hat{t}_1' \hat{t}_3' \frac{\hat{\sigma}_2 \hat{w}_{2123} + \hat{\sigma}_1 \hat{w}_{1223}}{(\hat{\sigma}_2^2 - \hat{\sigma}_1^2) \hat{\sigma}_2} \\ &+ \hat{t}_3' \hat{t}_1' \frac{\hat{\sigma}_2 \hat{w}_{2321} + \hat{\sigma}_2 \hat{w}_{2312}}{(\hat{\sigma}_2^2 - \hat{\sigma}_1^2) \hat{\sigma}_2} + \hat{t}_3' \hat{t}_3' \frac{\hat{w}_{2323}}{\hat{\sigma}_2^2} . \end{aligned}$$

We have

$$\hat{t}_1 = \begin{bmatrix} .508810 \\ .692941 \\ .510828 \end{bmatrix} \quad \hat{t}_3 = \begin{bmatrix} .649495 \\ -.698455 \\ .300528 \end{bmatrix} ,$$

$$\begin{aligned} \hat{\sigma}_1 &= 1 & \hat{\sigma}_2 &= .32669 \\ \hat{w}_{2123} &= -.040394 & \hat{v}_{4,4}^* &= .1995 \\ \hat{w}_{2321} &= -.040394 & \hat{w}_{1223} &= -.068420 \\ \hat{w}_{2323} &= .749728 & \hat{w}_{2312} &= -.068420 \end{aligned}$$

Proof. Since $n^{\frac{1}{2}}(\text{vec } \hat{\Sigma} - \text{vec } \Sigma)$ is asymptotically multivariate normal we can again apply the δ -method as long as we are careful to only use the functionally independent entries of $\hat{\Sigma}$. Let us work only with the entries r_{ij} and $\hat{\sigma}_{ij}$ with $i \geq j$. The distributions of r_{ij} ($i < j$) will follow from the fact that $r_{ij} = r_{ji}$. Using the δ -method again, $n^{\frac{1}{2}}(\ln r_{ij} - \ln \bar{r}_{ij})$ are asymptotically multivariate normal with means zero and covariances given by

$$\begin{aligned}
 & \text{asy. cov}(n^{\frac{1}{2}} \ln r_{ij}, n^{\frac{1}{2}} \ln r_{i',j'}) \\
 &= \sum_{\substack{g, h \\ g \geq h}} \sum_{\substack{g', h' \\ g' \geq h'}} \frac{\partial \ln \bar{r}_{ij}}{\partial \sigma_{gh}} \frac{\partial \ln \bar{r}_{i',j'}}{\partial \sigma_{g'h'}} \text{cov}(\hat{\sigma}_{gh}, \hat{\sigma}_{g'h'}) \\
 &= \sum_{\substack{g, h \\ g \geq h}} \sum_{\substack{g', h' \\ g' \geq h'}} \left(\frac{\delta_{ig} \delta_{jh}}{\sigma_{ij}} - \frac{1}{2} \frac{\delta_{ig} \delta_{ih}}{\sigma_{ii}} - \frac{1}{2} \frac{\delta_{jg} \delta_{jh}}{\sigma_{jj}} \right) \\
 &\quad \times \left(\frac{\delta_{i'g'} \delta_{j'h'}}{\sigma_{i'j'}} - \frac{1}{2} \frac{\delta_{i'g'} \delta_{i'h'}}{\sigma_{i'i'}} - \frac{1}{2} \frac{\delta_{j'g'} \delta_{j'h'}}{\sigma_{j'j'}} \right) \times (\sigma_{gh}, \sigma_{hg}, + \sigma_{gg}, \sigma_{hh}) \\
 &= \frac{\sigma_{ij}, \sigma_{ji}, + \sigma_{ii}, \sigma_{jj}}{\sigma_{ij} \sigma_{i'j'}} - \left[\frac{\sigma_{ii}, \sigma_{ji}}{\sigma_{ij} \sigma_{i'i'}} + \frac{\sigma_{ij}, \sigma_{jj}}{\sigma_{ij} \sigma_{j'j'}} + \frac{\sigma_{ij}, \sigma_{ii}}{\sigma_{ii} \sigma_{i'j'}} + \frac{\sigma_{jj}, \sigma_{ji}}{\sigma_{jj} \sigma_{i'j'}} \right] \\
 &\quad + \frac{1}{2} [\bar{r}_{ii}^2, + \bar{r}_{jj}^2, + \bar{r}_{ij}^2, + \bar{r}_{ji}^2,] \quad .
 \end{aligned}$$

Thus, $n^{\frac{1}{2}}(\text{vec } R - \text{vec } \bar{R})$ is asymptotically multivariate normal with variance-covariance matrix V_R , say. The entries of V_R can be found from

$$\begin{aligned}
 & \text{asy. cov}(n^{\frac{1}{2}} r_{ij}, n^{\frac{1}{2}} r_{i',j'}) = \bar{r}_{ij} \bar{r}_{i'j'} \left\{ \frac{\sigma_{ij}, \sigma_{ji}, + \sigma_{ii}, \sigma_{jj}}{\sigma_{ij} \sigma_{i'j'}} \right. \\
 &\quad \left. - \left[\frac{\sigma_{ii}, \sigma_{ji}}{\sigma_{ij} \sigma_{i'i'}} + \frac{\sigma_{ij}, \sigma_{jj}}{\sigma_{ij} \sigma_{j'j'}} + \frac{\sigma_{ij}, \sigma_{ii}}{\sigma_{ii} \sigma_{i'j'}} + \frac{\sigma_{jj}, \sigma_{ji}}{\sigma_{jj} \sigma_{i'j'}} \right] \right. \\
 &\quad \left. + \frac{1}{2} [\bar{r}_{ii}^2, + \bar{r}_{jj}^2, + \bar{r}_{ij}^2, + \bar{r}_{ji}^2,] \right\} \quad .
 \end{aligned}$$

We may now apply Theorem 1 to derive the asymptotic distributions of the eigenvectors and eigenvalues of R .

Theorem 5. Assumptions as in Lemma 4. Let the matrix \bar{R} have spectral decomposition,

$$\bar{R} = E \Lambda E' ,$$

where $EE' = E'E = I_p$ ($e_{ii} > 0$) and $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_p\}$ ($\lambda_1 > \lambda_2 > \dots > \lambda_p > 0$). Let R have a similar decomposition,

$$R = \hat{E} \hat{\Lambda} \hat{E}' ,$$

where $\hat{E} = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_p)$. Then the asymptotic distribution of the eigenvectors and eigenvalues is given by

$$\begin{aligned} n^{\frac{1}{2}}(\hat{\lambda} - \lambda) &\sim AN(0, V_{\hat{\lambda}}) \\ n^{\frac{1}{2}}(\hat{e}_k - e_k) &\sim AN(0, V_k) \quad k=1, 2, \dots, p , \end{aligned}$$

where

$$(V_{\hat{\lambda}})_{jk} = (e'_j \otimes e'_j) V_R (e_k \otimes e_k)$$

and

$$\text{asy. cov}(n^{\frac{1}{2}}\hat{e}_{jk}, n^{\frac{1}{2}}\hat{e}_{mn}) = \sum_{l \neq k} \sum_{l' \neq n} \frac{e_{jl} e_{ml'}}{(\lambda_k - \lambda_l)(\lambda_n - \lambda_{l'})} (e'_l \otimes e'_{k'}) V_R (e_l \otimes e_n) .$$

Proof. Use Theorem 1 and Lemma 4.

Remark. The above asymptotic results can be expected to take effect very slowly considering the slow approach to normality of r_{ij} .

6. SUMMARY

This paper has shown how results of Anderson [1963] can be applied to asymptotic eigenvector and eigenvalue analyses of matrices other than the sample variance-covariance matrix from a multivariate normal population. Results are derived for correspondence analysis and principal components analysis for correlation matrices. The results are considerably more complicated than the results for principal components of the sample variance-covariance matrix. In that case, the covariance structure of $\Gamma' \hat{\Sigma} \Gamma$ (from Theorem 1) takes a particularly simple form.

APPENDIX: PROOFS OF THEOREMS 1 AND 2

Both of the following proofs follow Anderson (1963) closely:

Proof of Theorem 1. Denote $T = \Gamma' \hat{\Sigma} \Gamma$ and let

$$U = n^{\frac{1}{2}}(\Gamma' \hat{\Sigma} \Gamma - \Lambda) = n^{\frac{1}{2}}(T - \Lambda) \quad .$$

Let the spectral decomposition of T be given by

$$T = E \hat{\Lambda} E' \quad , \quad (A1)$$

where $EE' = E'E = I_p$ and $e_{ii} > 0$. $\hat{\Lambda}$ appears since T has the same eigenvalues as $\hat{\Sigma}$ and the requirement that $e_{ii} > 0$ is to guarantee that the elements of E and $\hat{\Lambda}$ are uniquely specified. Also define

$$H = n^{\frac{1}{2}}(\hat{\Lambda} - \Lambda) \quad (A2)$$

$$f_{k\ell} = \begin{matrix} n^{\frac{1}{2}}e_{k\ell} & k \neq \ell \\ 0 & k = \ell \end{matrix} \quad . \quad (A3)$$

Using the convention that $\text{diag}\{A\}$ is an $n \times n$ diagonal matrix with the same diagonal elements as A , we can write,

$$\begin{aligned} \Lambda + n^{-\frac{1}{2}}U &= T \\ &= E \hat{\Lambda} E' \\ &= (\text{diag}\{E\} + E - \text{diag}\{E\}) \hat{\Lambda} (\text{diag}\{E\} + E - \text{diag}\{E\})' \\ &= (\text{diag}\{E\} + n^{-\frac{1}{2}}F) \hat{\Lambda} (\text{diag}\{E\} + n^{-\frac{1}{2}}F)' \\ &= \text{diag}\{E\} \hat{\Lambda} \text{diag}\{E\} + n^{-\frac{1}{2}}F \hat{\Lambda} \text{diag}\{E\} \\ &\quad + n^{-\frac{1}{2}}\text{diag}\{E\} \hat{\Lambda} F' + n^{-1}F \hat{\Lambda} F' \quad . \end{aligned}$$

Solving (A2) for $\hat{\Lambda}$ and inserting into the above yields the equation,

$$\begin{aligned}
 \Lambda + n^{-\frac{1}{2}}U &= \text{diag}\{E\} \Lambda \text{diag}\{E\} + n^{-\frac{1}{2}}\text{diag}\{E\}H\text{diag}\{E\} \\
 &+ n^{-\frac{1}{2}}F \Lambda \text{diag}\{E\} + n^{-1}FH\text{diag}\{E\} \\
 &+ n^{-\frac{1}{2}}\text{diag}\{E\} \Lambda F' + n^{-1}\text{diag}\{E\}HF' \\
 &+ n^{-1}F \Lambda F' + n^{-3/2}FHF' .
 \end{aligned} \tag{A4}$$

Since E is orthogonal,

$$\begin{aligned}
 I &= EE' \\
 &= (\text{diag}\{E\} + n^{-\frac{1}{2}}F)(\text{diag}\{E\} + n^{-\frac{1}{2}}F)' \\
 &= \text{diag}\{E\}\text{diag}\{E\} + n^{-\frac{1}{2}}F\text{diag}\{E\} \\
 &+ n^{-\frac{1}{2}}\text{diag}\{E\}F' + n^{-1}FF' .
 \end{aligned}$$

The on- and off-diagonal equations from the above matrix equation are

$$e_{kk}^2 = 1 - n^{-1}(FF')_{kk} \tag{A5}$$

$$0 = f_{kl}e_{ll} + f_{lk}e_{kk} + n^{-\frac{1}{2}}(FF')_{kl} \quad \begin{matrix} k, l = 1, 2, \dots, p \\ k \neq l \end{matrix} \tag{A6}$$

In matrix notation (A5) can be written

$$\text{diag}\{E\}\text{diag}\{E\} = I - n^{-1}\text{diag}\{FF'\} ,$$

which implies that

$$\text{diag}\{E\} \Lambda \text{diag}\{E\} = \Lambda - n^{-1}\text{diag}\{FF'\} \Lambda .$$

Using this in (A4) yields

$$\begin{aligned}
 \Lambda + n^{-\frac{1}{2}}U &= \Lambda - n^{-1}\text{diag}\{FF'\} \Lambda + n^{-\frac{1}{2}}\text{diag}\{E\}H\text{diag}\{E\} \\
 &+ n^{-\frac{1}{2}}F \Lambda \text{diag}\{E\} + n^{-1}FH\text{diag}\{E\} \\
 &+ n^{-\frac{1}{2}}\text{diag}\{E\} \Lambda F' + n^{-1}\text{diag}\{E\}HF' \\
 &+ n^{-1}F \Lambda F' + n^{-3/2}FHF' .
 \end{aligned}$$

Cancelling Λ from both sides, the on- and off-diagonal equations can be written as

$$u_{kk} = e_{kk}^2 h_{kk} + n^{-\frac{1}{2}} \text{ or smaller terms} \quad k=1,2,\dots,p \quad (A7)$$

$$u_{kl} = \lambda_k f_{lk} e_{kk} + \lambda_l f_{kl} e_{ll} + n^{-\frac{1}{2}} \text{ or smaller terms} \quad (A8)$$

$$k, l = 1, 2, \dots, p \quad k \neq l .$$

To derive the asymptotic distributions from (A7) and (A8) we need to show that $\text{diag}\{E\}$, H and F converge as U converges. This is proved in Anderson (1963, Section 7). Thus, in the limit,

$$\begin{aligned} u_{kk} &= e_{kk}^2 h_{kk} , \\ u_{kl} &= \lambda_k f_{lk} e_{kk} + \lambda_l f_{kl} e_{ll} , \\ 1 &= e_{kk}^2 , \\ 0 &= f_{kl} e_{ll} + f_{lk} e_{kk} , \end{aligned}$$

determine the asymptotic distribution of the h_{kk} , e_{kk} and f_{kk} . Thus, by Rubin's Theorem, the asymptotic distributions can be found as

$$e_{kk} \xrightarrow{\text{pr}} 1 , \quad (A9)$$

$$h_{kk} \text{ has the same asymptotic distribution as } u_{kk} , \quad (A10)$$

$$f_{lk} \text{ has the same asymptotic distribution as } -f_{kl} , \quad (A11)$$

$$f_{lk} \text{ has the same asymptotic distribution as } u_{lk}/(\lambda_k - \lambda_l) . \quad (A12)$$

Thus the asymptotic distributions translate directly from those of U . Since $n^{\frac{1}{2}}(\text{vec } \hat{\Sigma} - \text{vec } \Sigma) \sim AN(0, V)$, we have

$$n^{\frac{1}{2}}(\text{vec } U - \text{vec } \Lambda) \sim AN[0, (\Gamma' \otimes \Gamma)V(\Gamma \otimes \Gamma)] . \quad (A13)$$

Since $h_{kk} = n^{\frac{1}{2}}(\hat{\lambda}_k - \lambda_k)$, (A13) together with (A9) and (A10) prove (2.1). To prove (2.2) note that $\hat{\Gamma} = \Gamma E$, so

$$\hat{\gamma}_{\sim k} = \gamma_{\sim k} e_{kk} + n^{-\frac{1}{2}} \sum_{l \neq k} f_{lk} \gamma_{\sim l} .$$

Since $e_{kk} \xrightarrow{pr} 1$,

$$\begin{aligned} \text{asy. cov}(n^{\frac{1}{2}} \hat{\gamma}_{\sim j}, n^{\frac{1}{2}} \hat{\gamma}_{\sim k}) &= \text{asy. cov}\left(\sum_{l \neq j} f_{lj} \gamma_{\sim l}, \sum_{m \neq k} f_{mk} \gamma_{\sim m}\right) \\ &= \sum_{l \neq j} \sum_{m \neq k} \gamma_{\sim l} \gamma_{\sim m}' \text{asy. cov}(f_{lj}, f_{mk}) \\ &= \sum_{l \neq j} \sum_{m \neq k} \gamma_{\sim l} \gamma_{\sim m}' \text{cov}\left(\frac{u_{lj}}{\lambda_j - \lambda_l}, \frac{u_{mk}}{\lambda_k - \lambda_m}\right) \\ &= \sum_{l \neq j} \sum_{m \neq k} \gamma_{\sim l} \gamma_{\sim m}' \frac{(\gamma_{\sim l} \otimes \gamma_{\sim j}') V(\gamma_{\sim m} \otimes \gamma_{\sim k})}{(\lambda_j - \lambda_l)(\lambda_k - \lambda_m)} . \end{aligned}$$

Asymptotic normality follows from the asymptotic normality of the u_{lk} .

Proof of Theorem 2

Denote $R = S' \hat{B} T$ and let

$$Z = n^{\frac{1}{2}}[S' \hat{B} T - (\Sigma \ 0)] = n^{\frac{1}{2}}[R - (\Sigma \ 0)] . \quad (\text{A14})$$

Let the singular value decomposition of R be given by

$$R = E(\hat{\Sigma} \ 0)G' . \quad (\text{A15})$$

$\hat{\Sigma}$ appears since R has the same singular values as \hat{B} . Also define

$$\begin{aligned} K &= n^{\frac{1}{2}}(\hat{\Sigma} - \Sigma) , \\ F &= n^{\frac{1}{2}}(E - \text{diag}\{E\}) , \\ H &= n^{\frac{1}{2}}(G - \text{diag}\{G\}) . \end{aligned} \quad (\text{A16})$$

Using (A14) and (A15) we can write

$$\begin{aligned}
 (\Sigma \ O) - n^{-\frac{1}{2}}Z &= R \\
 &= E(\hat{\Sigma} \ O)G' \\
 &= (\text{diag}\{E\} + E - \text{diag}\{E\})(\hat{\Sigma} \ O)(\text{diag}\{G\} + G - \text{diag}\{G\}) \\
 &= (\text{diag}\{E\} + n^{-\frac{1}{2}}F)(\hat{\Sigma} \ O)(\text{diag}\{G\} + n^{-\frac{1}{2}}H) \\
 &= \text{diag}\{E\}(\hat{\Sigma} \ O)\text{diag}\{G\} + n^{\frac{1}{2}}F(\hat{\Sigma} \ O)\text{diag}\{G\} \\
 &\quad + n^{-\frac{1}{2}}\text{diag}\{E\}(\hat{\Sigma} \ O)H' + n^{-1}F(\hat{\Sigma} \ O)H' \quad .
 \end{aligned}$$

Solving (A16) for $\hat{\Sigma}$ and inserting in the above equation yields the equation,

$$\begin{aligned}
 (\Sigma \ O) + n^{-\frac{1}{2}}Z &= \text{diag}\{E\}(\Sigma \ O)\text{diag}\{G\} + n^{-\frac{1}{2}}\text{diag}\{E\}(K \ O)\text{diag}\{G\} \\
 &\quad + n^{-\frac{1}{2}}F(\Sigma \ O)\text{diag}\{G\} + n^{-1}F(K \ O)\text{diag}\{G\} \\
 &\quad + n^{-\frac{1}{2}}\text{diag}\{E\}(\Sigma \ O)H' + n^{-1}\text{diag}\{E\}(K \ O)H' \\
 &\quad + n^{-1}F(\Sigma \ O)H' + n^{-3/2}F(K \ O)H' \quad . \tag{A17}
 \end{aligned}$$

Since E and G are orthogonal we have

$$\text{diag}\{E\}\text{diag}\{E\} = I - n^{-1}\text{diag}\{FF'\}$$

and

$$\text{diag}\{G\}\text{diag}\{G\} = I - n^{-1}\text{diag}\{HH'\} \quad ,$$

which imply

$$\begin{aligned}
 e_{kk} &= 1 + n^{-1} \text{ or smaller terms} & k &= 1, 2, \dots, q \\
 g_{\ell\ell} &= 1 + n^{-1} \text{ or smaller terms} & \ell &= 1, 2, \dots, p \quad .
 \end{aligned}$$

Using these in (A17) yields the following

$$\begin{aligned}
 Z &= \text{diag}\{E\}(K \ O)\text{diag}\{G\} + F(\Sigma \ O)\text{diag}\{G\} + \text{diag}\{E\}(\Sigma \ O)H' \\
 &\quad + n^{-\frac{1}{2}} \text{ or smaller terms} \quad . \tag{A18}
 \end{aligned}$$

Since E is the diagonalizing matrix for RR' , convergence of R implies convergence of $\text{diag}\{E\}$, F and K by exactly the same argument as in the proof of

Theorem 1. Since $g_i = \operatorname{Re}_i \sigma_i^{-1}$ ($i=1,2,\dots,p$), this also implies convergence of g_{kk} and h_{kl} ($k,l=1,2,\dots,p$). Thus, in the limit, (A18) becomes

$$Z_{ii} = e_{ii} k_{ii} g_{ii}, \quad 1 = e_{ii} \quad \text{and} \quad 1 = g_{ii}.$$

By Rubin's Theorem, we obtain

$$e_{ii} \xrightarrow{\text{pr}} 1$$

$$g_{ii} \xrightarrow{\text{pr}} 1$$

k_{ii} has the same asymptotic distribution as Z_{ii} .

Since $n^{\frac{1}{2}} \operatorname{vec} Z \sim \operatorname{AN}[0, (T' \otimes S') Z (T \otimes S)]$, the result follows.

REFERENCES

- ANDERSON, T. W. (1963), "Asymptotic Theory for Principal Components Analysis," Annals of Mathematical Statistics, 34, 122-148.
- BISHOP, Y. M. M., FIENBERG, S. E., and HOLLAND, P. W. (1975), Discrete Multivariate Analysis, Cambridge, MA: MIT Press.
- HILL, M. O. (1973), "Reciprocal Averaging: An Eigenvector Method of Ordination," Journal of Ecology, 61, 237-251.
- HILL, M. O. (1974), "Correspondence Analysis: A Neglected Multivariate Method," Applied Statistics, 23, 340-354.
- HENDERSON, H. V., and SEARLE, S. R. (1979), "Vec and Vech Operators for Matrices, with Some Uses in Jacobians and Multivariate Statistics," Canadian Journal of Statistics, 7, 65-81.
- LANCASTER, H. O. (1958), "The Structure of Bivariate Distributions," Annals of Mathematical Statistics, 29, 719-736.
- MAGNUS, J. R., and NEUDECKER, H. (1979), "The Commutation Matrix: Some Properties and Applications," Annals of Statistics, 7, 381-394.
- MAUNG, K. (1941), "Measurement of Association in a Contingency Table with Special Reference to the Pigmentation of Hair and Eye Colours of Scottish School Children," Annals of Eugenics, London, 21, 189-223.
- O'NEILL, M. E. (1978a), "Asymptotic Distributions of the Canonical Correlations from Contingency Tables," Australian Journal of Statistics, 20, 75-82.
- O'NEILL, M. E. (1978b), "Distributional Expansions for Canonical Correlations from Contingency Tables," Journal of the Royal Statistical Society, Series B, 40, 303-312.
- O'NEILL, M. E. (1981), "A Note on Canonical Correlations from Contingency Tables," Australian Journal of Statistics, 23, 58-66.

SEBER, G. A. F. (1977), Linear Regression Analysis, New York: Wiley.

SWAN, J. M. A. (1970), "An Examination of Some Ordination Problems by Use of Simulated Vegetational Data," Ecology, 51, 89-102.

WHITTAKER, R. H. (1967), "Gradient Analysis of Vegetation," Biological Reviews, 42, 207-264.

Table 1

Eye color	<u>Hair color</u>			Total
	Fair	Medium	Dark	
Light to medium	5,335	8,634	3,217	17,186
Dark	<u>454</u>	<u>2,103</u>	<u>2,618</u>	<u>5,175</u>
Total	5,789	10,737	5,835	22,361